Richard Sprague

Metagenome Overview

MG-RAST ID 4633498.3 ⬇ Download 📊 Analyze 🔍 Search

Metagenome Name ssr_13603__R1__L001 **NCBI Proj**
PI **GOLD ID**
Organization **PubMed ID**
Visibility Private
Static Link http://metagenomics.anl.gov/linkin.cgi?metagenome=4633498.3

[Delete] [Share] [Edit Name] [Make Public]

METAGENOME SUMMARY

Dataset ssr_13603__R1__L001 was uploaded on 05/18/2015 and contains 43,577 sequences totaling 6,477,024 basepairs with an average length of 148 bps. The piechart below breaks down the uploaded sequences into 3 distinct categories.

2,769 sequences (6.4%) failed to pass the QC pipeline. Of the sequences that passed QC, 42,078 sequences (96.6%) contain ribosomal RNA genes. 0 (0.0%) of the sequences that passed QC have no rRNA genes.

The analysis results shown on this page are computed by MG-RAST. Please note that authors may upload data that they have published their own analysis for, in such cases comparison within the MG-RAST framework can not be done.

⬇ DOWNLOAD data and annotations

📊 ANALYZE annotations in detail.
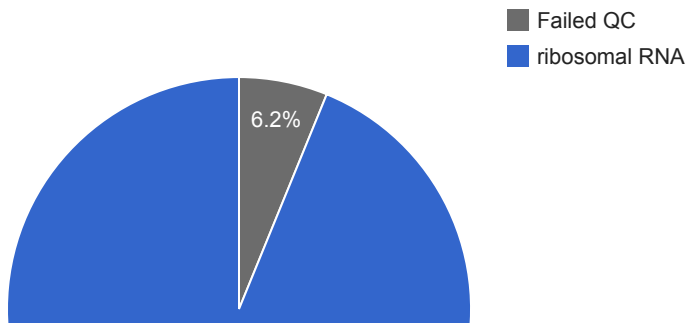
🔍 SEARCH through annotations.

Sequence Breakdown

■ Failed QC
■ ribosomal RNA

6.2%

TABLE OF CONTENTS

Note: Sequences containing multiple predicted features are only counted in one category.
Currently downloading of sequences via chart slices is not enabeled.

PROJECT INFORMATION

This dataset is part of project Sprague-Jan2015-13603.

There are 0 other metagenomes in this project

- » find metagenomes within this project
- » find metagenomes within this biome
- » find metagenomes within this country
- » find metagenomes within 10 | 30 | 100 kilometers

GSC MIxS INFO

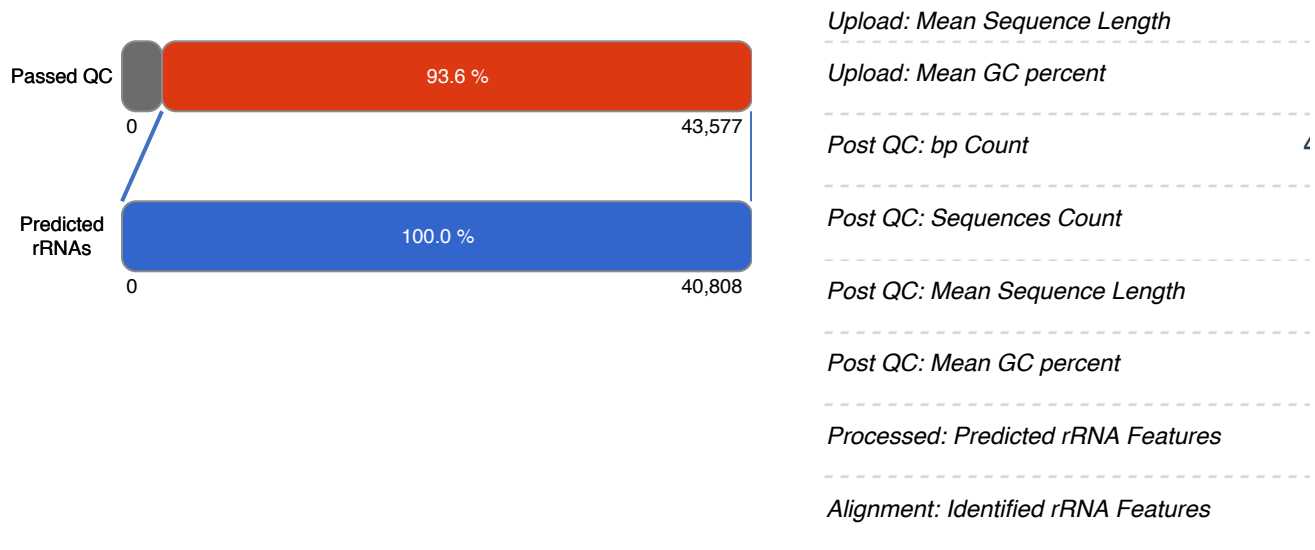|	
Investigation Type	**mimarks**
Project Name	**Sprague-Jan20**
Latitude and Longitude	
Country and/or Sea, Location	
Collection Date	
Environment (Biome)	
Environment (Feature)	
Environment (Material)	
Environmental Package	
Sequencing Method	
More Metadata	

ANALYSIS FLOWCHART

2,769 sequences failed quality control. Of the 40,808 sequences (totaling 4,753,862 bps) that passed quality control, 42,078 (103.1%) produced a total of 705 identified ribosomal RNAs.

ANALYSIS STATISTICS

Upload: bp Count	6
Upload: Sequences Count	

Passed QC

93.6 %

0   43,577

Predicted
rRNAs

100.0 %

0   40,808

Upload: Mean Sequence Length

Upload: Mean GC percent

Post QC: bp Count

Post QC: Sequences Count

Post QC: Mean Sequence Length

Post QC: Mean GC percent

Processed: Predicted rRNA Features

Alignment: Identified rRNA Features

DRISEE [?]

Duplicate Read Inferred Sequencing Error Estimation (Keegan et al., PLoS Computational Biology, 2012)

DRISEE could not produce a profile, this is an Amplicon dataset.

DRISEE is a tool that utilizes artificial duplicate reads (ADRs) to provide a platform independent assessment of sec sequencing data. DRISEE is designed to consider shotgun data. Currently, it is not appropriate for amplicon data.

Note that DRISEE is designed to examine sequencing error in raw whole genome shotgun sequence data. It assur been removed, but that the sequence data have not been modified in any additional way. (e.g.) Assembly or mergir DRISEE's ability to provide an accurate assessment of error by removing error before it is analyzed.

KMER PROFILES [?]   HIDE

Redraw the below plot using the following kmer-plot type: | kmer rank abundance ▼ | | 15-mer ▼ |

The kmer abundance spectra are tools to summarize the redundancy (repetitiveness) of sequence datasets by cou sequences.

The kmer spectrum plots the number of distinct N-bp sequences as a function of coverage level, placing low-cover repetitive sequences at right. The kmer rank abundance graph plots the kmer coverage as a function of abundance The ranked kmer consumed graph shows the fraction of the dataset that is explained by the most abundant kmers,

Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button | static |
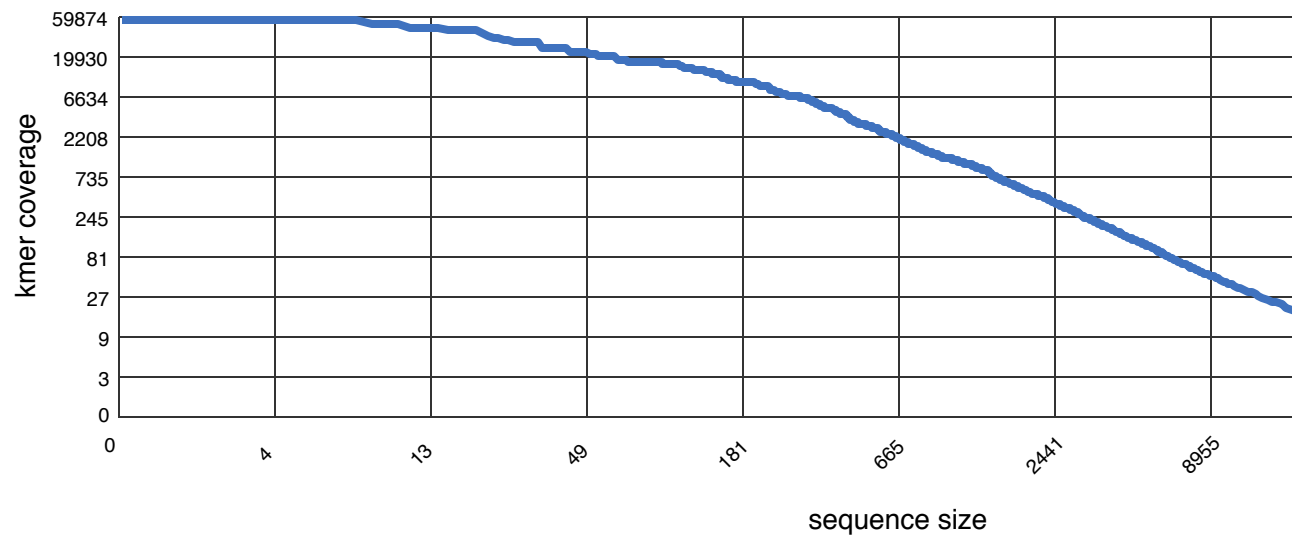
NUCLEOTIDE POSITION HISTOGRAM [?] HIDE

These graphs show the fraction of base pairs of each type (A, C, G, T, or ambiguous base "N") at each position sta 151 base pairs. Amplicon datasets should show consensus sequences; shotgun datasets should have roughly equ

Download chart data
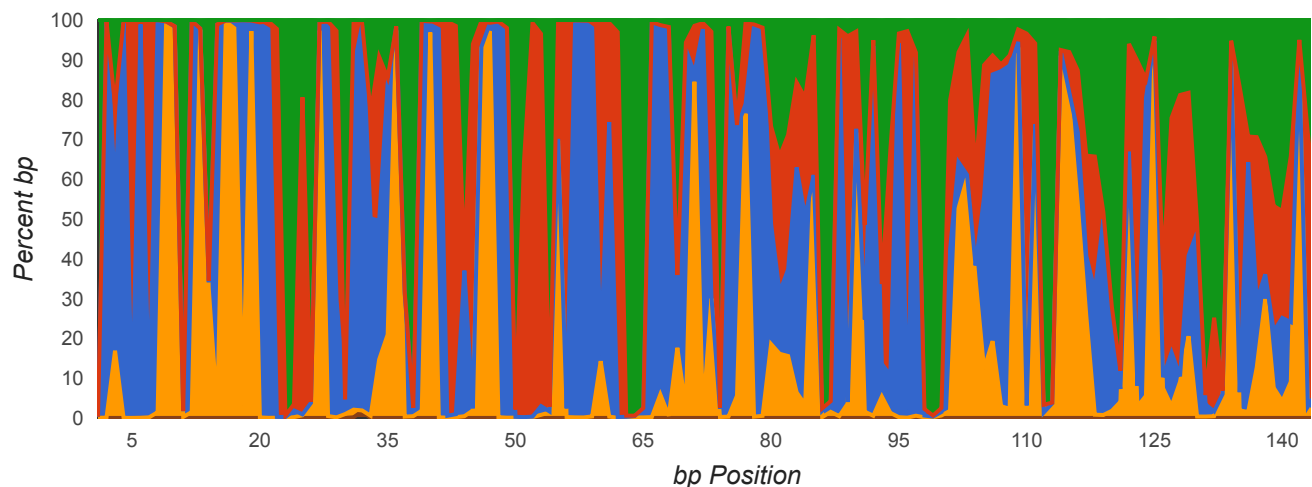
SOURCE HITS DISTRIBUTION [?] HIDE

42,078 (96.6%) of reads had similarity to ribosomal RNA genes.

The graph below displays the number of features in this dataset that were annotated by the different databases bel databases with functional hierarchy information, and ribosomal RNA databases. The bars representing annotated r databases have different numbers of hits, but can also have different types of annotation data.

There are 15,945,780 sequences in the M5NR protein database and 309,342 sequences in the M5RNA ribosomal the unique sequences from the below protein databases and the M5RNA ribosomal database contains all the uniqu

databases.

Download chart data
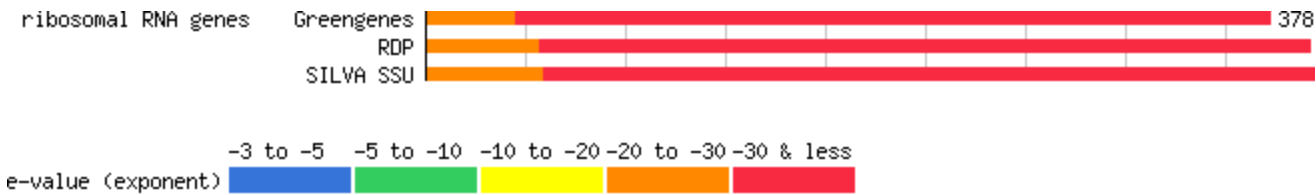
TAXONOMIC HITS DISTRIBUTION   HIDE

The pie charts below illustrate the distribution of taxonomic domains, phyla, and orders for the annotations. Each sl
proteins and ribosomal RNA genes annotated to the indicated taxonomic level. This information is based on all the
interactive Krona chart of the full taxonomy is also available
Click on a slice or legend to view all sequences annotated with the indicated taxonomic level in the analysis page.
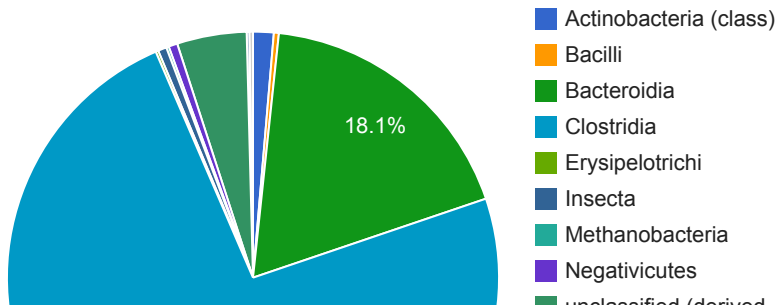
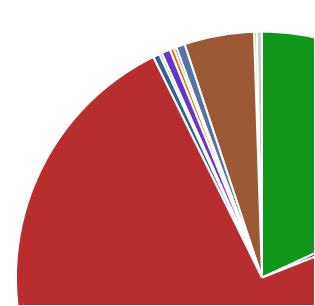View taxonomic interactive chart
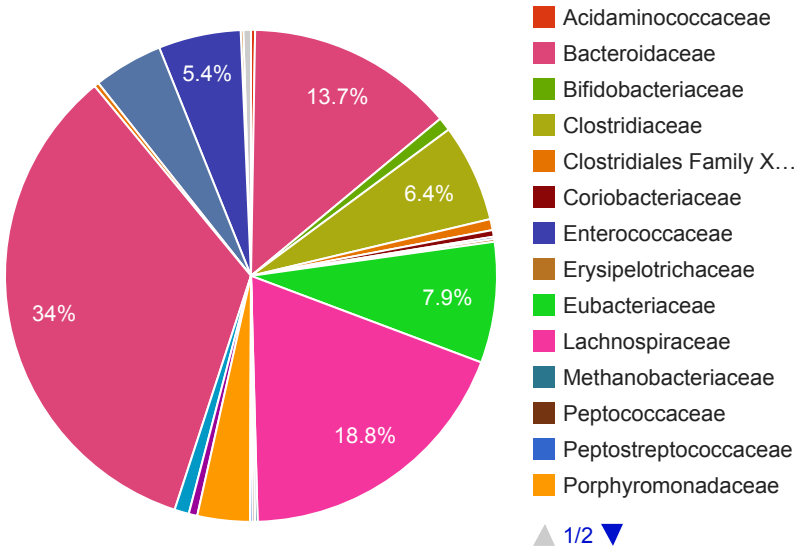
domain   Download chart data
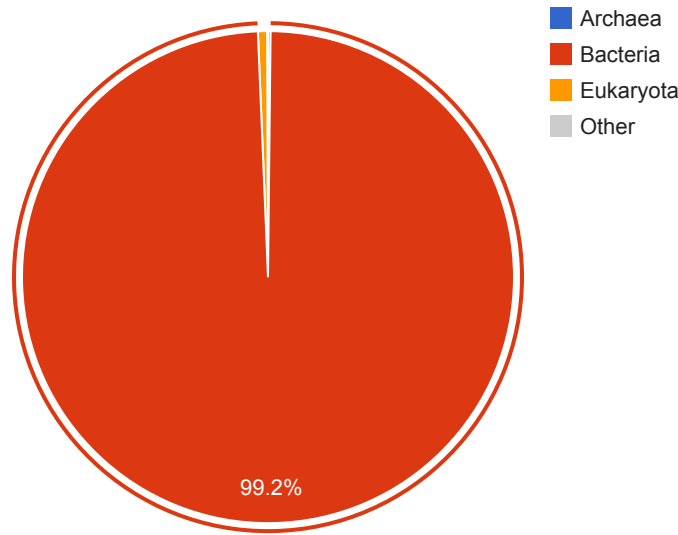
phylum   Download chart data
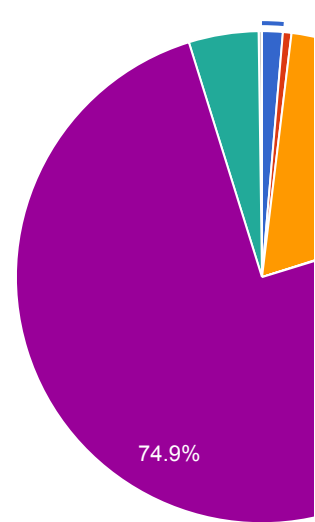
class   Download chart data

order   Download chart data

family   Download chart data

genus   Download chart data

- Acidaminococcaceae
- Bacteroidaceae
- Bifidobacteriaceae
- Clostridiaceae
- Clostridiales Family X…
- Coriobacteriaceae
- Enterococcaceae
- Erysipelotrichaceae
- Eubacteriaceae
- Lachnospiraceae
- Methanobacteriaceae
- Peptococcaceae
- Peptostreptococcaceae
- Porphyromonadaceae

△ 1/2 ▼
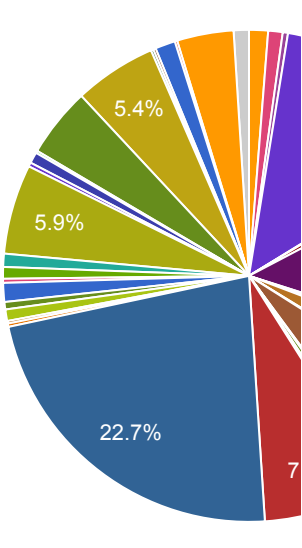
## RANK ABUNDANCE PLOT   HIDE

Redraw the below plot using the following taxonomic level: phylum ▼

The plot below shows the phylum abundances ordered from the most abundant to least abundant. Only the top 50 abundances of annotations in each phylum on a log scale.

The rank abundance curve is a tool for visually representing taxonomic richness and evenness.

Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button   static
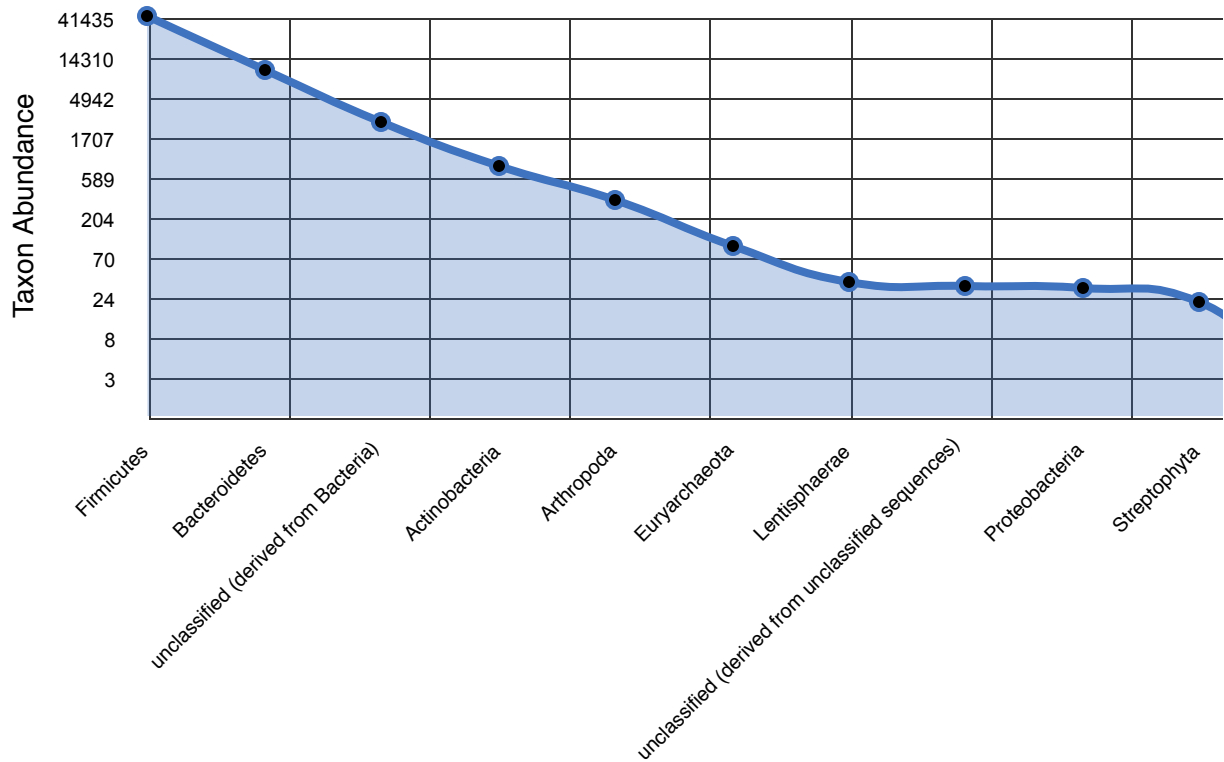
RAREFACTION CURVE   HIDE

The plot below shows the rarefaction curve of annotated species richness. This curve is a plot of the total number o
number of sequences sampled. On the left, a steep slope indicates that a large fraction of the species diversity rem
to the right, a reasonable number of individuals is sampled: more intensive sampling is likely to yield only few additi

Sampling curves generally rise very quickly at first and then level off towards an asymptote as fewer new species a
rarefaction curves are calculated from the table of species abundance. The curves represent the average number c
the the complete dataset.

Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button [static]
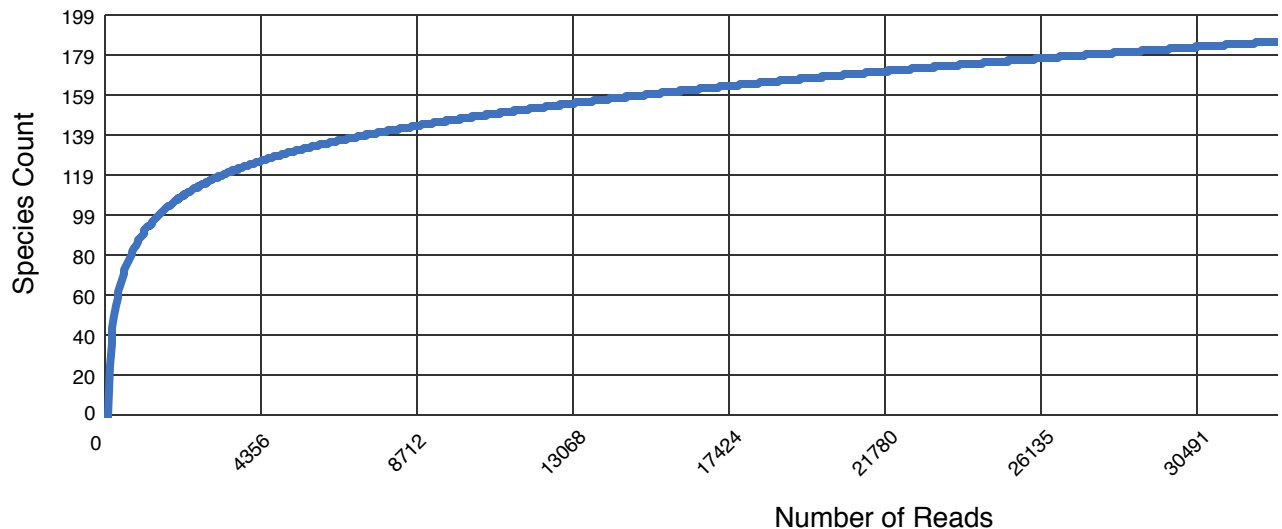
ALPHA DIVERSITY [?]   HIDE

α-Diversity = 31.330 species

The above image shows the range of α-diversity values in project Sprague-Jan2015-13603. The min, max, and me ranges (σ and 2σ) in different shades. The α-diversity of this metagenome is shown in red.

Alpha diversity summarizes the diversity of organisms in a sample with a single number. The alpha diversity of ann distribution of the species-level annotations.

Annotated species richness is the number of distinct species annotations in the combined MG-RAST dataset. Shar the logarithm of the relative abundances of annotated species. The species-level annotations are from all the annot

Download source data

SEQUENCE LENGTH HISTOGRAM   HIDE

The histograms below show the distribution of sequence lengths in basepairs for this metagenome. Each position r bp range.

The data used in these graphs are based on raw upload and post QC sequences.

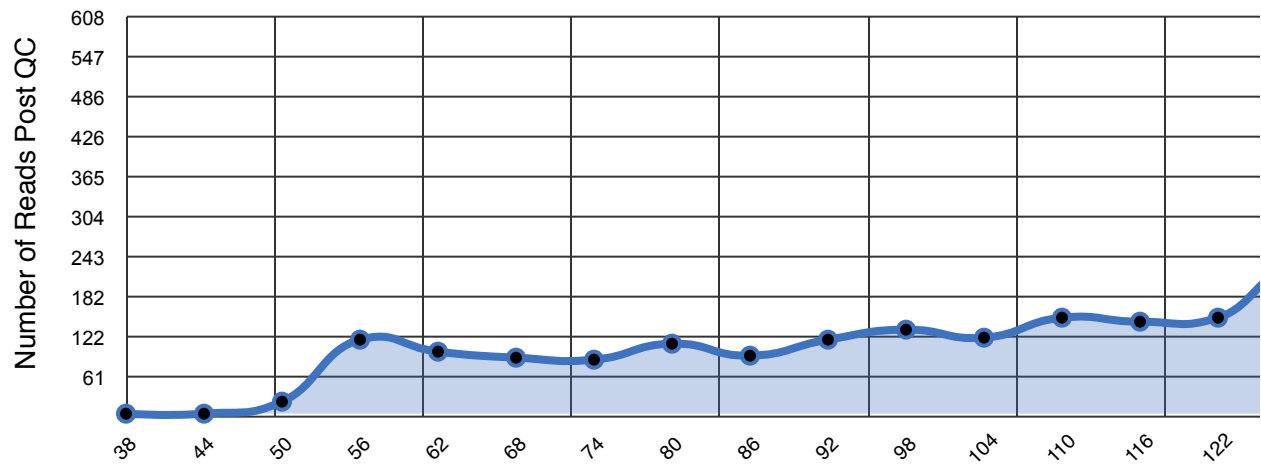Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button [static]
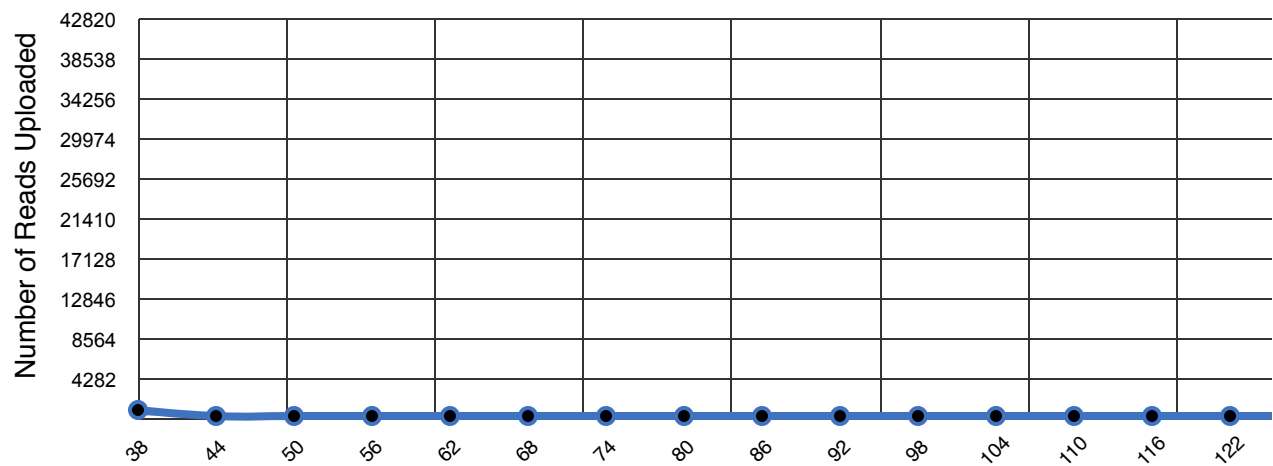
Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button [static]

SEQUENCE GC DISTRIBUTION   HIDE

The histograms below show the distribution of the GC percentage for this metagenome. Each position represents the range. The data used in these graphs is based on raw upload and post QC sequences.

Download chart data

The image is currently dynamic. To be able to right-click/save the image, please click the static button [static]

[Download chart data](#)

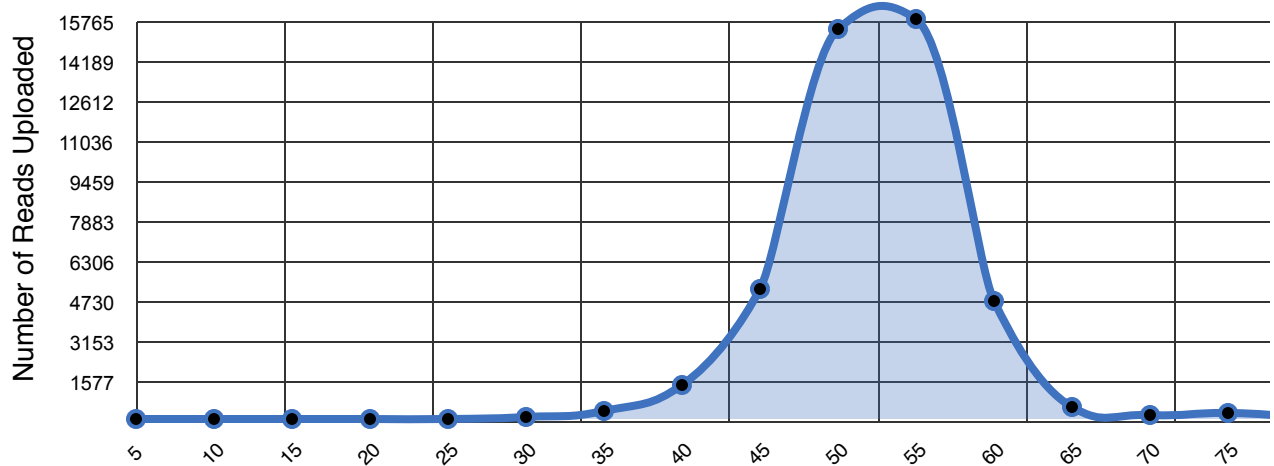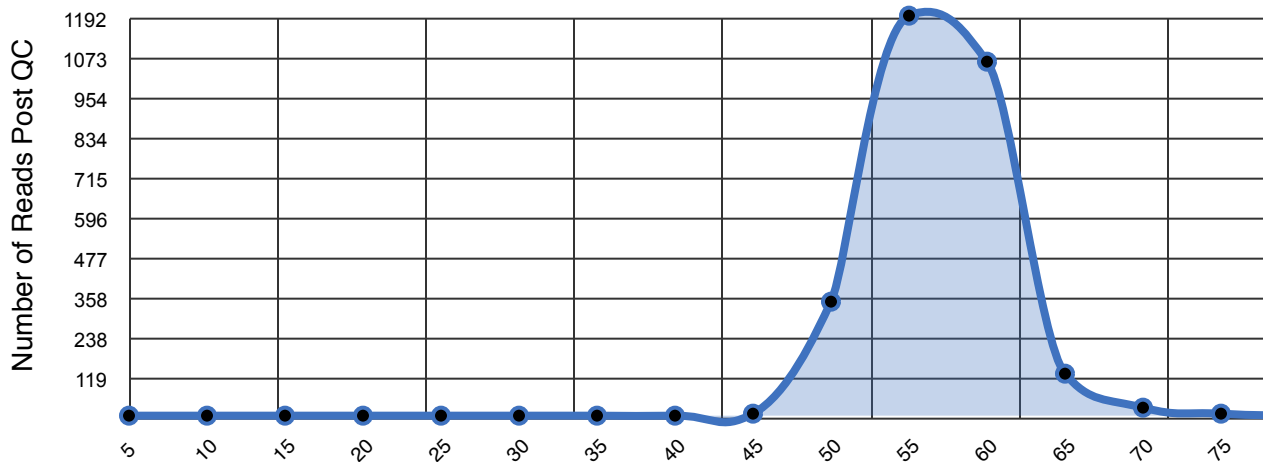The image is currently dynamic. To be able to right-click/save the image, please click the static button [static]

DOWNLOAD THIS METAGENOME

We provide download capabilities for the submitted sequences, metadata, and all files with results that are produce
download page for this metagenome. This includes fasta files with annotations using the M5NR.

We also provide access to the blat alignment summaries underlying our sequence analysis work on the download p

Please note: The graphs on this page allow downloading the underlying information as tables. The search results a
of sequences in an element to work with in the workbench feature on the analysis page.

ANALYZE THIS METAGENOME

The analysis page provides access to analysis and comparative tools including tables, bar charts, trees, principle c
(including FASTA and QIIME). The workbench feature allows sub-selections of data to be used e.g. select all E. col
present just in E. coli reads across multiple data sets.

SEARCH THIS METAGENOME

Below searches return all predicted functions or organisms that contain the input text.

[   ]   Search Functions

[   ]   Search Organisms

METADATA **[?]** SHOW